

(19)



KOREAN INTELLECTUAL PROPERTY OFFICE

KOREAN PATENT ABSTRACTS

(11)Publication number: 1020020019150 A
(43)Date of publication of application: 12.03.2002

(21)Application number: 1020000052274

(71)Applicant:

VOICE WEAR CO., LTD.

(22)Date of filing: 05.09.2000

(72)Inventor:

LEE, YUN GEUN

(51)Int. Cl

G10L 17/00

(54) METHOD OF GENERATING VOICE MODEL

(57) Abstract:

PURPOSE: A method of generating a voice model is provided to create a voice model in a tool kit using a variable word recognizer that easily generates and stores a voice model with respect to a specific object word.

CONSTITUTION: A word set including phonemes of a corresponding language is generated(S10). Voice data according to the word set is collected(S20). The voice data is labeled in phonemes(S30). The voice data is classified into short time periods, and spectrum of each period is analyzed to extract voice characteristic to represent each phoneme in rows of characteristic vector(S40). The same phonemes contained in the voice data are collected to perform HMM training(S50). HMMs of phonemes constructing a voice model to be generated are connected(S60).



COPYRIGHT KIPO 2002

Legal Status

Date of final disposal of an application (20030108)

Patent registration number (1003694780000)

Date of registration (20030113)

Date of opposition against the grant of a patent (00000000)

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(51) Int. Cl.
G10L 17/00

(11) 공개번호 특2002-0019150
(43) 공개일자 2002년03월12일

(21) 출원번호	10-2000-0052274
(22) 출원일자	2000년09월05일
(71) 출원인	(주) 보이스웨어 백종관
(72) 발명자	이윤근
(74) 대리인	서울특별시 동작구 흑석2동명수대현대0107-1403 최덕규

심사청구 : 있음

(54) 음성 모델의 생성 방법

요약

본 발명은 음성 모델을 생성하는 방법에 관한 것으로, 화자 독립 방식의 가변어 인식기를 포함하는 툴킷(tool kit)에서 은닉 마코프 모델의 음성 인식 기법을 이용하여 음성 모델을 생성하는 방법에 관한 것이다. 그 방법은 해당 언어의 음소가 포함되는 단어 셋(set)을 생성하고, 상기 단어 셋에 따른 음성 데이터를 수집하고, 상기 음성 데이터를 각각의 음소 단위로 레이블링(labeling)을 하고, 상기 음성 데이터를 짧은 시간 구간으로 분류하여 각 구간의 스펙트럼 분석을 하여 음성특징을 추출해서 각각의 음소를 특징벡터의 열로 나타내고, 상기 음성 데이터에 포함되어 있는 같은 음소를 모아 HMM훈련을 수행하고 그리고 생성하고자 하는 음성 모델을 구성하는 음소들의 상기 HMM을 연결하는 단계로 이루어진다.

도표도

도1

색인어

음성 인식, 음성 모델, 음소, 가변어 인식기, 툴킷, 화자 종속 방식, 화자 독립 방식, 은닉 마코프 모델

명세서

도면의 간단한 설명

제1도는 본 발명에 따라 가변어 인식기에서 음성 모델(HMM)을 생성하는 과정을 흐름도로 도시한 것이다.
제2도는 본 발명에 따라 가변어 인식기에서 음성 모델을 생성하는 과정을 개념적으로 도시한 것이다.
제3도는 음성모델을 저장하는 음성 인식용 칩을 블록도로 도시한 것이다.

발명의 상세한 설명

발명의 목적

본 발명이 속하는 기술분야 및 그 분야의 종래기술

발명의 분야

본 발명은 음성 모델을 생성하는 방법에 관한 것이다. 보다 구체적으로 본 발명은 화자 독립 방식의 가변어 인식기를 이용한 툴킷(tool kit)에서 원하는 단어 셋(set)에 대한 음성 모델을 은닉 마코프 모델의 음성 인식 기법을 이용하여 생성하는 방법에 관한 것이다. 또한 메모리에 상기한 방법으로 생성된 음성 모델을 저장하여 음성 인식 기능을 수행할 수 있도록 하는 칩을 제공하기 위한 것이다.

발명의 배경

음성 인식은 전화, 휴대폰 또는 마이크를 통하여 전달된 사람의 음성 특징을 추출하고 분석하여, 미리 입력된 음성 모델에서 가장 근접한 결과를 찾아내는 최첨단 기술이다. 음성 인식용 칩은 일본의 오오키, 미국의 세서리 등 여러 회사에서 이미 제품화되어 생산하고 있다. 현재 제품화되어 사용하는 음성 인식용 칩은 음성 인식을 적용할 대상 단어가 선정되면 상기 대상 단어에 해당하는 음성 모델을 생성하여 칩의 메모리에 탑재하게 되어 있다.

상기 과정을 좀 더 상세히 설명하면 다수의 사람으로부터 음성을 채취하여 저장하고, 각 회사에서 제공하는 툴킷(tool kit)을 이용하여 상기 채취된 음성을 음성모델로 변환하여 주고, 그리고 상기 음성 모델을 장치의 메모리에 탑재하는 절차로 이루어진다.

상기와 같은 절차로 이루어지는 종래의 음성 인식용 칩은 화자 종속 방식의 음성 인식 형태로 음성 모델을 생성하기 때문에 인식 대상 단어가 변경되는 경우에 있어서는 새롭게 음성 모델을 만들어 칩의 메모리에 탑재하여야 한다. 따라서 시간도 많이 소비되어질 뿐만 아니라 비용도 많이 드는 단점이 있다. 대상 단어에 제한을 받지 않기 위해, 가변어 인식기를 칩에 탑재할 수도 있으나, 이런 경우에는 알고리즘이 복잡해지며 메모리도 많이 소요되므로 장치의 가격이 비싸지는 단점이 있다.

이에 본 발명자는 화자 독립 방식의 가변 인식을 하는 가변어 인식기를 이용한 툴킷에서 일정한 대상 단어에 대한 음성 모델을 용이하게 생성할 수 있는 방법을 개발하게 된 것이다.

본 발명이 이루고자 하는 기술적 과제

본 발명의 목적은 상기에 제시된 문제점들을 해소할 수 있도록 가변어 인식 알고리즘을 이용해서 일정한 대상 단어에 대한 음성모델을 손쉽게 생성하여 저장할 수 있는 가변어 인식을 하는 가변어 인식기를 이용한 툴킷에서 음성 모델을 생성하는 방법을 제공하기 위한 것이다.

본 발명의 다른 목적은 음성 인식을 위한 음성 모델을 생성하는데 있어 다수의 음성을 채취하여 녹음하여 음성 모델로 모델링하는 절차를 포함한 번잡한 프로세스를 용이하게 수행할 수 있도록 한 가변어 인식을 하는 가변어 인식기를 이용하여 음성 모델을 생성하는 방법을 제공하기 위한 것이다.

본 발명의 또 다른 목적은 종래 화자 종속 방식의 음성 인식 형태로 음성 모델을 생성하기 때문에 인식 대상 단어가 변경되는 경우에 있어서는 새롭게 음성모델을 만들어야 하는 문제점을 극복할 수 있도록 한 가변어 인식을 하는 가변어 인식기를 이용하여 음성 모델을 생성하는 방법을 제공하기 위한 것이다.

본 발명의 상기 및 기타의 목적들은 하기 설명되는 본 발명에 의하여 모두 달성될 수 있다.

본 발명의 구성 및 작용

본 발명은 해당 언어의 음소가 포함되는 단어셋을 생성하고, 상기 단어셋에 따른 음성 데이터를 수집하고, 상기 음성 데이터를 각각의 음소 단위로 레이블링(labeling)을 하고, 상기 음성 데이터를 짧은 시간 구간으로 분류하여 각 구간의 스펙트럼 분석을 하여 음성특징을 추출해서 각각의 음소를 특징벡터의 열로 나타내고, 상기 음성 데이터에 포함되어 있는 같은 음소를 모아 HMM훈련을 수행하고 그리고 생성하고자 하는 음성 모델을 구성하는 음소들의 상기 HMM을 연결하는 단계로 이루어진다.

본 발명에서는 인식 대상 단어의 음성 모델을 가변어 인식을 하는 가변어 인식기를 이용한 툴킷에서 생성한다. 상기 툴킷은 가변어 인식기와 음소 단위의 음성 모델 데이터베이스를 포함하고 있다. 상기 가변어 인식기는 은닉 마코프 모델(HMM : Hidden Markov Model) 방식에 의해 인식 대상 단어를 모델링하여 화자 독립 방식의 음성 인식 기능을 제공하는 것으로, 해당 기술분야에서 널리 사용되고 있는 것일 뿐만 아니라 해당 기술 분야에서 통상의 지식을 가진 자라면 누구나 용이하게 실시할 수 있으므로 상세한 설명은 생략한다. 상기 은닉 마코프 모델은 각 단어의 발음상에 존재하는 각종 변화 요인들을 마코프 모델에 근거를 둔 통계적 기법으로 묘사하여 패턴 유사도를 측정하는 방법으로, 훈련 과정에서 마코프 프로세스에서의 상태전이 확률 및 출력 심볼 관찰 확률을 추정한다. 인식과정에서는 이들 확률로부터 비터비 디코딩(Viterbi decoding)에 의해 인식단어를 결정하게 된다. 상기과 같은 음성 인식 기법 역시 해당 분야에서 통상의 기술을 가진 자에 의해서 용이하게 실시되어질 수 있으므로 상세한 설명은 생략하기로 한다. 덧붙여서 본 발명은 상기과 같은 종래의 음성 인식 기법 및 관련 기술을 이용하여 가변어 인식을 수행하는 화자 독립 방식의 가변어 인식기를 포함하는 툴킷(tool kit)에서 음성 모델을 보다 용이하게 생성하는 방법을 제공하고자 하는 것이다.

이하에서는 첨부하는 도면을 가지고 상기과 같은 종래의 툴킷 및 음성 인식 기법을 이용하여 음성모델을 생성하는 방법을 설명하기로 한다. 도1은 가변어 인식기에서 음성 모델(HMM)을 생성하는 과정을 도시한 것이다.

먼저 인식 대상 단어의 모든 음소가 포함되어 있는 단어(PBW : Phoneme Balanced Word) 셋을 만든다(S10).

상기 PBW 셋을 다수의 사람들이 발성하여 만들어지는 음성 데이터를 수집한다(S20).

상기 수집된 음성 데이터를 각 음소 단위로 레이블링을 한다(S30).

상기 음성 데이터를 짧은 시간구간(frame)으로 나누어 각 구간마다 스펙트럼 분석을 하여 음성특징(feature)을 추출하면 상기 음성 데이터의 각각의 음소는 특징 벡터의 열로 나타내어진다. 각각의 다른

발음의 음소들은 독특한 스펙트럼 특징을 가지고 있으며 동일한 음소들은 비슷한 스펙트럼 특징을 가지고 있으므로 이러한 특징은 음성 인식에 이용된다(S40).

상기 음성 데이터에 포함되어 있는 같은 음소를 모아 HMM 훈련을 거친다(S50). 상기 HMM은 여러 개의 상태(state)로 구성되어 있는 확률적 모델로써 각 음소마다 다른 HMM이 발생된다. HMM의 각 상태(state)는 시간 정보를 가지며 각 상태(state)마다 특정한 특징(feature) 벡터의 출력 확률을 갖는다. 따라서, 각 음소의 HMM에는 그 음소의 시간에 따른 주파수 특성이 나타나 있다. 본 단계에서는 최적의 모델링을 하기 위해 각 파라미터를 조정하기 위해서 바움-웰치(Baum-Welch) 알고리즘을 이용한다.

마지막으로 구하고자 하는 음성 모델을 구성하는 음소들의 상기 HMM을 연결시켜 음성모델을 생성한다(S60).

상기와 같이 이루어지는 본 발명에 따른 음성모델을 생성하는 방법을 요약하여 설명하면, 도2에 개념적으로 도시한 바와 같이 음성 인식 대상 단어가 선정되면 톨킷에 대상 단어 리스트를 입력하게 된다. 상기 톨킷을 이루고 있는 가변어 인식기는 각각의 대상 단어를 음소 단위로 분해한 후 상기 대상 단어의 음소와 대응하는 데이터베이스에 저장되어 있는 각각의 음소 모델들을 연결시켜 음성 모델을 생성하게 되는 것이다.

상기와 같이 음성 모델을 생성한 후 상기 음성 모델을 칩의 메모리에 저장을 함으로써 음성 인식용 칩은 구성이 된다. 이하에서는 상기와 같은 절차로 생성된 음성 모델을 저장하는 음성 인식용 칩에 대해 설명하기로 한다.

도3은 음성 모델을 저장하는 음성 인식용 칩의 구성을 블록도로 도시한 것이다. 도시한 바와 같이 음성 인식용 칩은 음성을 입력하는 입력부(101), 상기 입력부에서 입력된 아날로그 신호인 음성을 디지털로 변환하거나 디지털 신호를 아날로그 신호로 변환하는 변환부(102), 상기 변환부에서 입력된 음성 디지털 신호와 톨킷에서 생성된 음성 모델을 비교, 연산하는 연산부(103), 톨킷에서 생성된 음성모델 및 칩의 동작에 필요한 프로그램을 저장하는 저장부(104)로 구성된다. 또한 메모리를 추가하는 경우에는 데이터버스를 통하여 외부메모리를 추가할 수도 있다.

상기 외부 메모리는 플래쉬(flash) 메모리, 롬 또는 램으로 구성되어 메모리의 추가를 손쉽게 할 수 있다.

상기 입력부(101)는 화자가 입으로 발생하는 음파를 전기적 신호로 바꾸어 준다.

변환부(102)는 상기 입력부(101)에서 입력된 음성 신호를 디지털신호로 변환(coder)하고 디지털신호를 역으로 아날로그 신호로 변환하는(decoder) 변, 복조기 또는 부호, 복호기를 말한다.

상기 연산부(103)는 상기 변환부(102)에서 화자의 음성 아날로그 신호를 A/D(아날로그/디지털)변환하여 얻어진 디지털 데이터에 대수적인 연산을 해 필터링이나 스펙트럼 분석 등의 신호처리를 수행하며 기본적으로 아날로그 신호의 실시간 디지털 처리를 목적으로 한다. 상기 연산부(103)는 디지털신호처리의 기본 조작으로 필터링, 푸리에(fourier)변환, 상관함수의 산출, 부호화, 부호 변복조, 미분, 적분, 적응신호처리 등을 수행하게 된다.

상기 저장부(104)는 음성인식을 위한 프로그램을 저장하는 프로그램 메모리.

상기 저장부(104)와 연산부(103)와 입·출력에서 전달되는 데이터 교환의 효율성을 높이기 위하여 이들 사이에서 교환되는 데이터가 잠시동안 저장될 수 있도록 버퍼 및 작업용 메모리 등으로 구성된다. 또한 상기 저장부(104)에서 상기에서 설명한 가변어 인식기를 포함하는 톨킷에서 생성한 음성모델을 탑재하고 있다. 만일 상기와 같이 구성되는 음성인식용 칩을 원-칩(one-chip)형태로 구성하는 경우에는 상기 음성 모델을 저장부(104)의 롬(rom)에 마스킹하게 된다.

이하에서는 상기와 같이 구성되는 음성인식 용 칩의 동작을 설명하기로 한다.

선행적으로 상기에서 설명한 가변어 인식기를 포함하는 톨킷에서 음성모델을 생성한 후 상기 음성모델을 저장부(104)의 플래시 메모리 또는 롬에 저장을 한다.

입력부(101)는 화자가 발음하는 음성을 변환부(102)에 입력을 한다. 상기 화자가 발음하는 음파를 상기 변환부(102)에서는 디지털 신호인 음성 데이터로 변환하게 된다. 그러면 연산부(103)는 저장부(104)에서 음성인식 알고리즘을 불러 들여 상기 변환부(102)에서 입력된 음성데이터에서 음성 특징 벡터열을 추출하고 이 벡터열을 이용하여 상기 저장부(104)에 저장되어 있는 음성 모델의 HMM 과 입력된 음성 특징 벡터열과의 유사도(likelihood)를 측정하여 가장 유사도가 높은 것을 인식 결과로 결정한다. 일반적으로 상기 와 같이 음성 인식을 하는데 있어서는 관측열이 주어졌을 때 최적의 상태를 선택하기 위해서 비터비(viterbi) 알고리즘을 이용한다.

발명의 효과

본 발명에 따라 이루어지는 음성 모델을 생성하는 방법은 가변어 인식 알고리즘을 이용해서 일정한 대상 단어에 대한 음성모델을 손쉽게 생성하여 저장할 수 있으며, 음성 인식을 위한 음성 모델을 생성하는데 있어 다수의 음성을 채취하여 녹음하여 음성 모델로 모델링하는 절차를 포함한 번잡한 프로세스를 용이하게 수행할 수가 있다. 또한 인식 대상 단어가 변경되는 경우에 있어서도 용이하게 음성모델을 생성할 수가 있으며 인식 대상 단어를 손쉽게 업데이트할 수가 있는 것이다.

본 발명의 단순한 변형 내지 변경은 이 분야의 통상의 지식을 가진 자에 의하여 용이하게 이용될 수 있으며, 이러한 변형이나 변경은 모두 본 발명의 영역에 포함되는 것으로 볼 수 있다.

(57) 청구의 범위

청구항 1

해당 언어의 음소가 포함되는 단어셋을 생성하고;

상기 단어셋에 따른 음성 데이터를 수집하고;

상기 음성 데이터를 각각의 음소 단위로 레이블링을 하고;

상기 음성 데이터를 짧은 시간 구간으로 분류하며 각 구간의 스펙트럼을 분석하여 음성특징을 추출해서 각각의 음소를 특징벡터의 열로 나타내고;

상기 음성 데이터에 포함되어 있는 같은 음소를 모아 HMM훈련을 수행하고; 그리고

생성하고자 하는 음성 모델을 구성하는 음소들의 상기 HMM을 연결하는;

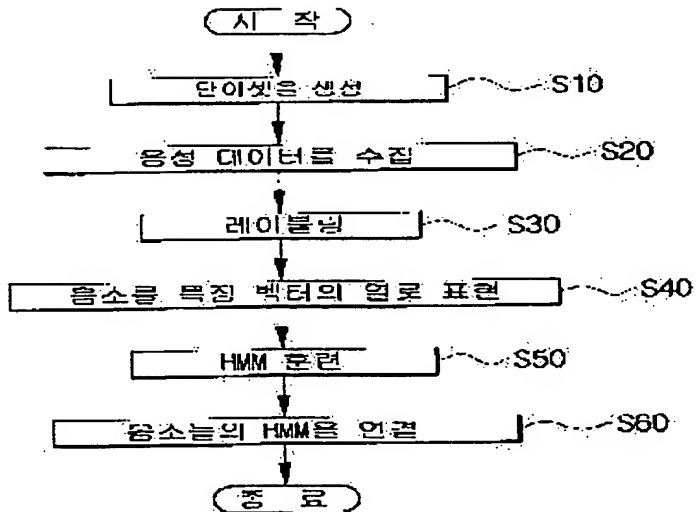
단계로 이루어지는 가변 인식을 하는 화자 독립 방식의 가변어 인식기를 포함하는 툴킷(tool kit)에서 음성 모델을 생성하는 음성 모델의 생성 방법.

청구항 2

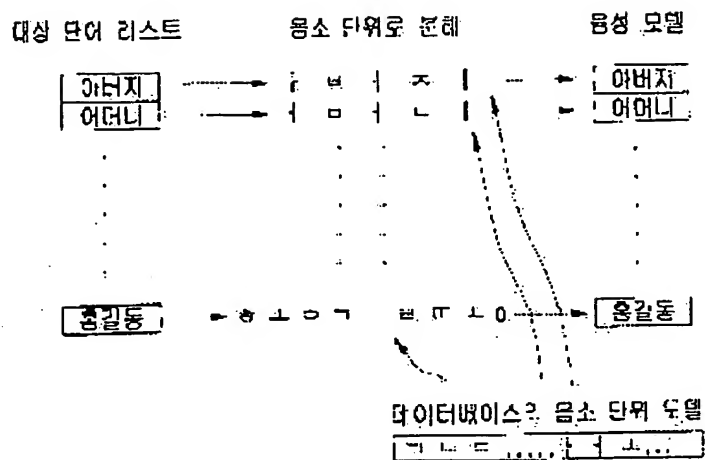
제1항의 방법에 의해 가변 인식을 하는 화자 독립 방식의 가변어 인식기를 포함하는 툴킷에서 생성된 음성 모델을 저장하고 있는 음성 인식용 칩.

도면

도면1



도면2



도면3

